

(19)日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11)特許出願公表番号

特表平11-504733

(43)公表日 平成11年(1999)4月27日

(51)Int.Cl.⁶
G 1 0 L 9/14
H 0 3 M 7/30
H 0 4 B 14/04

識別記号

F I
G 1 0 L 9/14
H 0 3 M 7/30
H 0 4 B 14/04

J
B
Z

審査請求 未請求 予備審査請求 未請求(全 34 頁)

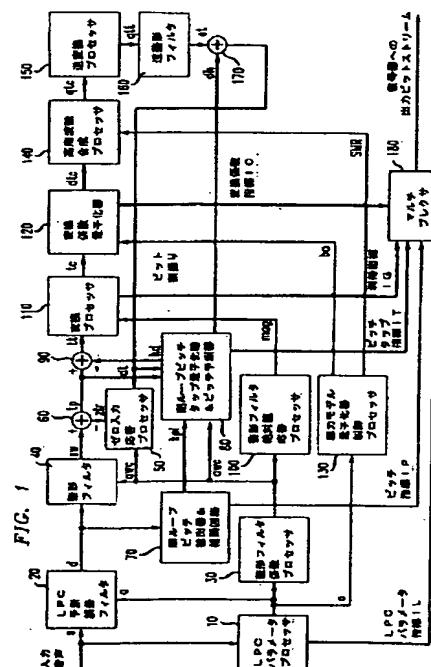
(21)出願番号 特願平9-530382
(86) (22)出願日 平成9年(1997)2月26日
(85)翻訳文提出日 平成9年(1997)10月27日
(86)国際出願番号 PCT/US97/02898
(87)国際公開番号 WO97/31367
(87)国際公開日 平成9年(1997)8月28日
(31)優先権主張番号 60/012,296
(32)優先日 1996年2月26日
(33)優先権主張国 米国(US)
(81)指定国 EP(AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), CA, JP, MX, US

(71)出願人 エイ・ティ・アンド・ティ・コーポレーション
アメリカ合衆国 10013-2412 ニューヨーク ニューヨーク アヴェニュー オブ
ジ アメリカズ 32
(72)発明者 チン ジュイン ウェイ
アメリカ合衆国 ニュージャージー州 ネ
シャニック ステーション ロングフィー
ルド ドライブ 68
(74)代理人 弁理士 吉田 研二 (外2名)

(54)【発明の名称】 聴覚モデルによる量子化を伴う予測残余信号の変形符号化による多段音声符号器

(57)【要約】

「変換予測符号化」又はTPCと呼ばれる音声圧縮システムは、1サンプルあたり1つ又は2つの16又は32kb/sのビットという標準ビットレートでサンプリングする16kHzでの7kHz帯域の音声の符号化を行う。システムは、冗長性を除去するため、短期及び長期の予測を用いる。予測残余は、(60)からの時間領域データ及び(100)からのパラメータ入力を受諾した後(110)により図に示されたとおり周波数領域で変換され符号化され、かくして聴覚知覚のためのスペクトルが補正される。TPC符号器は(70)により示されているような開ループ量子化のみを用い、従って低い複雑性をもつ。音声品質は32kb/sで透明であり、24kb/sで非常に優れ、16kb/sで許容できるものである。



【特許請求の範囲】

1. 音声信号のフレームを符号化する方法において、
予測残余信号を生成するべく線形予測フィルタを使用して音声信号から短期相
関関係を除去するステップと、
予測残余信号に基づき音声信号の開ループピッチ周期推定を決定するステップ
と、
予測残余信号の量子化されたバージョンに基づいてフレームの2つ以上のサブ
フレームについてのピッチフィルタタップの重みを決定するステップと、
開ループピッチ周期推定、2つ以上のサブフレームについてのピッチフィルタ
タップの重み、及び予測残余信号に基づいて、ピッチ予測残余信号を形成するス
テップと、
ピッチ予測残余信号を量子化するステップと、
を含んで成る方法。

【発明の詳細な説明】

聴覚モデルによる量子化を伴う予測残余信号の変形符号化による多段音声符号器 発明の分野

本発明は、例えば音声信号といったオーディオ信号の予測符号化システムを用いた圧縮（符号化）に関する。

発明の背景

信号圧縮に関する文献で教示されているように、音声及び音楽の波形は、非常に異なる符号化技術によって符号化される。16 k b／秒以下で符号化する電話帯域幅（3・4 kHz）の音声といったような音声符号化は、時間領域予測符号器により主に行われてきた。これらの符号器は、符号化すべき音声波形を予測するのに音声生成モデルを使用する。このとき、原信号内の冗長性を低減するため実際の（もとの）（符号化すべき）波形から予測された波形が減算される。信号冗長性の減少は、符号化利得をもたらす。このような予測音声符号器の例としては、音声信号圧縮の技術において全て良く知られたものである適応予測符号化、マルチパルス線形予測符号化及び符号励起線形予測（CELP）符号化がある。

一方、64 k b／秒以上の広帯域（0～20 kHz）音楽符号化は、周波数領域変換又はサブバンド符号器によって主に行われてきた。これらの音楽符号器は、根本的に上述の音声符号器と非常に異なっている。この差異は、音楽ソースが、音声ソースと異なり直ちに予測できるようにするには変動が大きすぎるものであるという事実に起因する。その結果、音楽ソースのモデルは一般に音楽符号化には使用されない。その代り、音楽符号器は、信号のうち知覚的に関連する部分のみを符号化するために、精巧な人間の聴覚モデルを使用する。すなわち、一般に音声生成モデルを用いる音声符号器とは異なり、音楽符号器は、符号化利得を得るべく聴力（受音）モデルを利用する。

音声符号器においては、符号化すべき音楽の雑音マスキング能力を見極めるために、聴力モデルが使用される。「雑音マスキング能力」という語は、リスナー

が雑音に気づかうことなく音楽信号内にどれほどの量子化雑音を導入できるかを言う。この雑音マスキング能力はこのとき、量子化器分解能（例えば量子化器ス

ステップサイズ)をセットするのに用いられる。一般に、音楽は、それが「音調様」になればなるほど、マスキング量子化雑音において粗末なものとなり、従って所要ステップサイズは小さくなり、その逆もあてはまる。より小さなステップサイズは、より小さい符号化利得に対応し、逆も又言える。このような音楽符号器の例としては、A T & T の知覚オーディオ符号 (P A C) 及び I S O M P E G オーディオ符号化規格が含まれる。

電話帯域幅音声符号化と広帯域音楽符号化の間には、音声信号が 16 kHz でサンプリングされ 7 kHz の帯域幅をもつ広帯域音声符号化が存在する。7 kHz の広帯域音声のもつ利点は、結果として得られる音声品質が電話帯域幅音声よりもはるかに優れており、しかも符号化に必要となるビットレートは 20 kHz のオーディオ信号よりもはるかに低い。以前に提案されたこれらの広帯域音声符号器の中には、時間領域予測符号化を使用するものもあれば、周波数領域変換又はサブバンド符号化を使用するものもあり、さらには、時間領域と周波数領域の両技術の混合を使用するものもある。

広帯域であれその他のものであれ、予測音声符号化に知覚基準を含み入れることは、複数の合成音声信号候補の中から最良の合成音声信号を選択するという場合での知覚重みづけフィルタの使用に限られてきた。例えば、Atal et al に対する米国特許 Re. 32580 号を参照のこと。このようなフィルタは、符号化プロセスで雑音を減少させるのに有用である、あるタイプの雑音整形を達成する。ある既知の符号器は、知覚重みづけフィルタの形成において知覚モデルを利用することにより、この技術を改善する試みを行なっている。

発明の要約

上述の努力にも関わらず、既知の音声又はオーディオ符号器のいずれも、信号雑音マスキング能力の分析に従って量子化器分解能をセットするための聴力モデルと、信号予測の目的の音声生成モデルの両方を利用してはいない。

一方、本発明は、雑音に対する人間の聴覚感度モデルで決定された雑音マスキング信号に基づいて信号を量子化する量子化プロセスと予測符号化システムを組合わせている。予測符号化システムの出力はかくして、オーディオ知覚モデルに

従って決定された雑音マスキング信号の一関数である分解能（例えば、一様スカラ量子化器におけるステップサイズ、又はベクトル量子化器においてベクトルを識別するのに用いられるビット数）をもつ量子化器で量子化される。

本発明によれば、音声情報を表わす信号の推定（又は予測）を表わす1つの信号が生成される。「音声情報を表わす原信号」という語は、音声自体のみならず音声符号化システム内に一般に見られる音声信号派生物（例えば線形予測及びピッチ予測残余信号）をも意味するほどに充分広義のものである。このとき推定信号は原信号と比較されて、これらの比較された信号の間の差異を表わす信号を形成する、比較された信号の間の差を表わすこの信号、次に、人間のオーディオ知覚モデルよって生成される知覚雑音マスキング信号に従って量子化される。

「変換予測符号化」又はT P Cと呼ばれる本発明の実施形態は、16～32 k b／秒のターゲットビットレートで7 kHzもの広帯域音声を符号化する。その名が示すように、T P Cは単一の符号器の中で、変換符号化と予測符号化の技術を組合せている。より特定的に言うと、符号器は、入力音声波形から冗長性を除去するのに線形予測を使用し、次に結果として得た予測残余を符号化するため変換符号化技術を使用する。変換された予測残余は、可聴なものを符号化し非可聴なものを廃棄するべく、聴覚知覚モデルの形で表現された人間の聴知覚における知識に基づいて量子化される。

実施形態の1つの重要な特長は、信号の知覚雑音マスキング能力（例えば「ちょうど認識可能なひずみ」の知覚的なしきい値）が決定されその後のビット割振りが行なわれる方法に関する。従来の音楽符号器において行なわれているように、未量子化入力信号を用いて知覚しきい値を決定するのではなく、この実施形態の雑音マスキングしきい値及びビット割振りは、量子化された合成フィルタ（この実施形態では量子化されたL P C合成フィルタ）の周波数応答に基づいて決定される。この特長は、復号器が受理され符号化された広帯域音声情報を復号するのに必要とされる知覚しきい値及びビット割振り処理を複製するために、符号器から復号器までビット割振り信号を伝達する必要がない、という利点をシステムに

提供する。その代り、その他の目的のために伝達されつつある合成フィルタ係数が、ビットレートを節約するために開発利用される。

実施形態のもう1つの重要な特長は、T P C符号器が符号器周波数の間でいかにビットを割振りするか、又割振られたビットに基づいて復号器が量子化された出力信号をいかにして生成するかに関するものである。或る種の状況下では、T P C符号器は、オーディオ帯域の一部分にのみビットを割振る（例えば、0～4 k H zの間の係数に対してのみビットを割振ることができる）。4 k H zと7 k H zの間の係数を表わすのにいかなるビットも割振られず、かくして復号器はこの周波数範囲内でいかなる係数も得ない。このような状況は、T P C符号器が例えば16 k b／秒といったひじょうに低いビットレートで作動しなければならない場合に発生する。4 k H z及び7 k H zの周波数範囲内で符号化された信号を表わすビットを全くもたないにもかかわらず、復号器はそれでも、広帯域応答を提供しなければならない場合この範囲内の信号を合成しなくてはならない。実施形態のこの特長に従うと、復号器は、その他の利用可能な情報すなわち（L P Cパラメータから得られた）信号スペクトルの推定とその範囲内の周波数での雑音マスキングしきい値の比率とに基づいてこの周波数範囲内の係数信号を生成（すなわち合成）する。この技術によって、復号器は、全帯域について音声信号係数を伝達する必要なく広帯域応答を提供することができる。

広帯域音声符号器の混在的利用分野としては、I S D Nテレビ会議又はオーディオ会議、マルチメディアオーディオ、「ハイファイ」電話技術及び28.8 k b／秒以上でモデムを用いたダイアル呼出し回線上での同時ボイス&データ（S V D）が含まれる。

図面の簡単な説明

図1は、本発明の符号器実施形態の図を示す。

図2は、本発明の復号器実施形態の図を示す。

図3は、図1のL P Cパラメータプロセッサの詳細なブロックダイヤグラム図を示す。

詳細な説明

A. 実施形態の概要

説明を明確にするため、図に示す本発明の実施形態は、個々の機能ブロック（「プロセッサ」と標識付けされている機能ブロックを含む）を含んで成るものとして示されている。これらのブロックが表わす機能は、ソフトウェアを実行することのできるハードウェアを含む（ただしこれに限られるわけではない）共用か又は専用のハードウェアの使用を通して提供できるものである。例えば、図1～4に示されているプロセッサの機能は、単一の共用プロセッサによって提供されてもよい。（「プロセッサ」という語の使用は、ソフトウェアを実行できるハードウェアを排他的に指すものとみなされてはならない）。

図に示す実施形態には、A T & T D S P 1 6 又は D S P 3 2 C といったようなデジタル信号プロセッサ (D S P) ハードウェア、以下に論述するオペレーションを実行するソフトウェアを記憶するための読み取り専用メモリ (R O M) 及び D S P 結果を記憶するためのランダムアクセスメモリ (R A M) が含まれてもよい。同様に、超大規模集積回路 (V L S I) ハードウェアの実施形態ならびに汎用 D S P 回路と組合わせたカスタム V L S I 回路を具備してもよい。

本発明に従うと、デジタル入力音声サンプルのシーケンスは、フレームと呼ばれる連続した 2 0 m s のブロックに区分され、各々のフレームはさらに、各々 4 m s の 5 つの等しいサブフレームに細分される。広帯域音声信号にとって一般的であるように 1 6 k H z のサンプリングレートを仮定すると、これは、3 2 0 サンプルというフレームサイズ及び 6 4 サンプルというサブフレームサイズに対応する。T P C 音声符号器は入力音声信号をフレーム毎に緩衝するとともに処理し、各々のフレーム内で、いくつかの符号化オペレーションがサブフレーム毎に実行される。

図1は、本発明のT P C 音声符号器の実施形態の一例を示す。図1に示された実施形態を参照されたい。2 0 m s のフレーム毎に一回、L P C パラメータプロセッサ 1 0 は、入力音声信号 S から線スペクトル対 (L S P) パラメータを導出し、このような L S P パラメータを量子化し、各々の 4 m s のサブフレームについてこれらを補間し、次に各サブフレームについて L P C 予測器係数アレイ a に

変換する。短期冗長性は、LPC予測誤差フィルタ20により入力音声信号sから除去される。結果として得られたLPC予測残余信号dは、なお、有聲音内のピッチ周期性に起因する幾つかの長期冗長性をもつ。整形フィルタ係数プロセッサ30は、量子化されたLPCフィルタ係数aから整形フィルタ係数awcを導出する。整形フィルタ40はLPC予測残余信号dをろ過して知覚的に重みづけされた音声信号swを生成する。ゼロ入力応答プロセッサ50は、整形フィルタのゼロ入力応答zirを計算する。減算ユニット60はswからzirを減算して、ピッチ予測用の標的信号tpを得る。

開ループピッチ抽出器及び補間回路70は、各々の20msフレームについてピッチ周期を抽出するため、LPC予測残余dを使用し、次に各々の4msサブフレームについて、補間されたピッチ周期kpiを計算する。閉ループピッチタップ量子化器及びピッチ予測器80は、この補間されたピッチ周期kpiを用いて、ピッチタップの候補セットのコードブックから3つのピッチ予測器タップを1セット選択する。この選択は、以前に量子化されたLPC残余信号dtが対応する3-タップピッチ合成フィルタによってろ過され、次にゼロ初期メモリを備える整形フィルタによってろ過された時点で、出力信号hdが平均二乗誤差(MSE)センス内の標的信号tpに最も近いところにあるような形で行なわれる。減算ユニット90は、tpからhdを減算して、変換符号化のための標的信号ttを得る。

整形フィルタ絶対応答プロセッサ100は、整形フィルタの周波数応答の絶対値である信号magを計算する。変換プロセッサ110は、信号ttについて、高速フーリエ変換(FFT)といったような線形変換を実行する。次に、このプロセッサは、3つの異なる周波数帯域について計算された利得値の量子化バージョン及びmagを用いて変換係数を正規化する。結果は、正規化された変換係数信号tcである。このとき、変換係数量子化器120は、異なる周波数での変換係数の時間変動する知覚的大きさに従って、聴覚モデル量子化器制御プロセッサ130によって決定される適応ビット割振り信号baを用いて、信号tcを量子化する。

16kb/sといった低いビットレートで、プロセッサ130は、周波数帯域

(0～4 kHz) の下半分にのみビットを割振る。この場合、高周波数合成プロセッサ140が、高周波数帯域(4～8 kHz)内で変換係数を合成し、これらを量子化された低周波数変換係数信号 $d_t c$ と組合わせて、最終的な量子化された全帯域変換係数信号 $q_t c$ を生成する。24又は32kb/秒といったより高いビットレートでは、周波数帯域全体の中の各々の変換係数は、適応ビット割振りプロセスにおいてビットを受理することが許されるものの、場合によっては、利用可能なビットの不足に起因して全くビットを受理しない可能性もある。この場合、高周波数合成プロセッサ140は単に、ビットを全く受理しない4～8 kHz帯域内の周波数を検出し、適応変換符号器において標準的に見られる「うずまき」タイプのひずみを避けるため低レベル雑音でこのような「スペクトルホール」を充てんする。

逆変換プロセッサ150は、量子化された変換係数信号 $q_t c$ を取り上げ、変換プロセッサ110において利用された線形変換の逆オペレーションである線形変換（ここでの我々の特定の実施形態では逆FFT）を適用する。この結果、変換符号化のための標的信号である t_t の量子化バージョンである時間領域信号 $q_t t$ がもたらされる。このとき、逆整形フィルタ160は、 $q_t t$ をろ過して、量子化された励起信号 e_t を得る。加算器170は、ブロック80内のピッチ予測器により生成された信号 d_h （これはLPC予測残余 d のピッチ予測バージョンである）に対して e_t を加算する。結果として得られた信号 d_t は、LPC予測残余 d の量子化されたバージョンである。これは、ゼロ入力応答プロセッサ50内部の整形フィルタのフィルタメモリ及びブロック80内のピッチ予測器のメモリを更新する。こうして信号ループが完成する。

LPC予測器パラメータ(IL)、ピッチ予測器パラメータ(IP及びIT)、変換利得レベル(IG)及び量子化変換係数(IC)を表わすコードブック指標は、マルチプレクサ180によって1つのビットストリーム内に多重化され、チャンネル上で復号器まで伝送される。チャンネルは、無線チャンネル、コンピュータ及びデータネットワーク、電話回線網を含む適当なあらゆる通信チャンネルを含んでいてよく、さらに、固体メモリ（例えば半導体メモリ）、光学メモリシステム（例えばCD-ROM）、磁気メモリ（例えばディスクメモリ）などを

含

んでいてよい。

図2は、本発明のT P C音声復号器実施形態を示す。デマルチフルクサ200は、コードブック指標IL, IP, IT, IG及びICを分離する。ピッチ復号器及び補間回路205は、IPを復号し、補間されたピッチ周期k_{pi}を計算する。ピッチタップ復号器及びピッチ予測器210は、ピッチ予測器タップアレイbを得るべくITを復号し、又、信号d_h又はLPC予測残余dのピッチ予測されたバージョンを計算する。LPCパラメータ復号器及び補間回路215はILを復号し、次に補間されたLPCフィルタ係数アレイaを計算する。ブロック220から255までは、量子化されたLPC残余信号d_tを生成するべく、図1内のその対応構成と全く同じオペレーションを実行する。長期ポストフィルタ260は、d_t内のピッチ周期性を増強し、その出力としてろ過されたバージョンf_{d_t}を生成する。この信号はLPC合成フィルタ265内を通過させられ、結果として得られた信号s_tはさらに短期ポストフィルタ270によってろ過され、このフィルタ270は最終的ろ過済み出力音声信号f_{s_t}を生成する。

複雑性を低く保つため、T P Cはできるかぎり、開ループ量子化を利用する。開ループ量子化というのは、量子化器が、出力音声品質に対する影響とは無関係に、量子化されていないパラメータとその量子化バージョンの間の差異を最小限にしようと試みることを意味する。これは、ピッチ予測器、利得そして励起が通常閉ループ量子化されるC E L P符号器とは対照的である。符号器パラメータの閉ループ量子化においては、量子化器コードブック探索は、最後の再構築された出力音声内のひずみを最小限にしようとする。当然のことながら、これは一般により良い出力音声品質を導くが、その代価は、さらに高いコードブック探索の複雑性である。

本発明では、T P C符号器は、3ピッチ予測器タップについてのみ閉ループ量子化を使用する。量子化された励起信号e_tを導く量子化オペレーションは、基本的に開ループ量子化に類似しているが、出力音声に対する効果は、閉ループ量子化のものに近い。このアプローチの精神はLefebvre et al., 「変換符号化励

起 (T C X) を用いた広帯域オーディオ信号の高品質符号化」、Proc. IEEE International Conf. Acoustics, Speech, Signal Processing, 1994, pp. I

—193～I-196によりT C X符号器の中で用いられたアプローチに似ている。例えば、T C X符号器の中には本発明の特長としては、整形フィルタ絶対値応答による変換係数の正規化、聴力モデルにより制御される適応ビット割振りそして高周波数合成及び雑音充てん手順がある。

B. 符号器実施形態

1. L P C 解析及び予測

L P C パラメータプロセッサ 1 0 の詳細なブロックダイヤグラムが図 3 に示されている。プロセッサ 1 0 は、窓かけ及び自己相関プロセッサ 3 1 0 ; スペクトル平滑化及び白雑音補正プロセッサ 3 1 5 ; レヴィンソン-ダービン巡回形プロセッサ 3 2 0 ; 帯域幅拡張プロセッサ 3 2 5 ; L P C - L S P 変換プロセッサ 3 3 0 及び L P C パワースペクトルプロセッサ 3 3 5 ; L S P 量子化器 3 4 0 ; L S P ソーティングプロセッサ 3 4 5 ; L S P 補間プロセッサ 3 5 0 及び L S P - L P C 変換プロセッサ 3 5 5 を含む。

窓かけ及び自己相関プロセッサ 3 1 0 は、L P C 係数生成プロセスを開始する。プロセッサ 3 1 0 は、以下で論述する通り L P C 係数がその後計算される基となる自己相関係数 r を 2 0 m s 毎に一回の従来の要領で生成する。Rabiner L.R. et al., 音声信号のデジタル処理, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, 1978 (Rabiner et al.) を参照のこと。L P C フレームサイズは 2 0 m s (又は、1 6 k H z のサンプリングレートで 3 2 0 の音声サンプル) である。各々の 2 0 m s フレームはさらに、各々 4 m s (又は 6 4 サンプル) の長さの 5 つのサブフレームに分割される。L P C 解析プロセッサは、従来の要領で、現行フレームの最後の 4 m s のサブフレームにセンタリングされる 2 4 m s のハミング窓を使用する。

潜在的な不良条件づけを軽減するため、いくつかの従来の信号条件づけ技術が利用される。スペクトル平滑化技術 (S S T) 及び白雑音補正技術は、L P C 解析の前にスペクトル平滑化及び白雑音補正プロセッサ 3 1 5 によって適用される

。当該技術分野では周知のものである SST (Tohkura, Y. et al., 「PARC OR 音声解析－合成におけるスペクトル平滑化技術」 I E E E Trans, Acoust, S

peech, Signal Processing, A S S P - 2 6 : 5 8 9 - 5 9 6, 1978年12月 (Tohkura et al.)) にはが、40 Hz の標準偏差でのガウス分布の確率密度関数 (p d f) に対応するが、ガウシアン窓に、計算された自己相関係数アレイ (プロセッサ 310 から) を掛け合わせることが言及されている。同じく従来通りのものである (Chen, J - H, 「16 k b i t / 秒での頑強な低遅延 C E L P 音声符号器」, Proc I E E E, Global Comm. Conf. p p. 1237 ~ 1241, Dallas, TX, 1989年11月) 白雑音補正は、ゼローラグ自己相関係数 (すなわちエネルギー項) を 0.001 % だけ増大させる。

このとき、プロセッサ 315 により生成された係数は、レビンソンダービン巡回形プロセッサ 320 に提供され、このプロセッサ 320 は、従来の要領で $i = 1, 2, \dots, 16$ についての 16 の L P C 係数 a_i (L P C 予測誤差フィルタ 20 の次数は 16 である) を生成する。

帯域幅拡張プロセッサ 325 は、さらなる信号条件付けのため各々の a_i に g_i という係数を乗じる (なおここで、 $g = 0.994$ である)。これは、30 Hz の帯域幅拡張に対応する (Tohkura et al.)

このような帯域幅拡張の後、L P C 予測器係数は、従来の要領で L P C - L S P 変換プロセッサ 330 により線スペクトル対 (L S P) 係数に変換される。Soong, F.K. et al., 「線スペクトル対 (L S P) 及び音声データ圧縮」 Proc. I E E E Int. Conf. Acoust. Speech, Signal Processing, p p 1.10.1 - 1.10.4, 1984年3月 (Soong et al.) を参照されたい。なおこの文献はここにあたかも完全に記述されているかのごとく参考として含まれるものである。

次に、結果として得られた L S P 係数を量子化するための L S P 量子化器 340 によりベクトル量子化 (V Q) が提供される。プロセッサ 240 により利用される特定の V Q 技術は、ここにあたかも完全に記述されているかのごとく参考として含まれる, Paliwal, K. K. et al., 「24 ビット / フレームでの L P C パ

ラメータの効率の良いベクトル量子化」, Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing pp 661~664, Toronto, Canada, 199

1年5月(Paliwal et al)の中で提案されている分割VQに類似している。16

次元LSPベクトルは、低周波数端部から計数して2, 2, 2, 2, 2, 3, 3

の次元をもつ7つのさらに小さいベクトルに分割される。7つのサブベクトルの各々は、7ビットに量子化される（すなわち128のコードベクトルのVQコードブックを用いて）。かくして、各々長さ7ビットの7つのコードブック指標IL(1)~IL(7)が存在し、合計でLPCパラメータ量子化において1フレームにつき49ビットが使用される。これらの49個のビットは、サイド情報として復号器に対する伝送のためマルチプレクサ180に提供される。

プロセッサ340は、Paliwal et alに記述されている通り、従来の重みづけされた平均二乗誤差(WMSE)ひずみ測定を用いて、VQコードブックを通してその探索を実行する。LPCパワースペクトルプロセッサ335は、このWMSEひずみ測定における重みを計算するために使用される。プロセッサ340で使用されるコードブックは、当該技術分野において周知の従来のコードブック生成技術を用いて設計される。従来のMSEひずみ測定は同様に、出力音声品質における重大な劣化なく符号器の複雑性を低減するべく、WMSE測定の代りに使用することもできる。

通常、LSP係数は、単調に増大する。しかしながら、量子化は、この秩序の破断を結果としてもたらす可能性がある。この破断の結果、復号器内のLPC合成フィルタは不安定なものとなる。この問題を回避するため、LSPソーティングプロセッサ345は、量子化されたLSP係数をソートして、単調に増加する秩序を復元し安定性を確保する。

量子化されたLSP係数は、現フレームの最後のサブフレーム内で使用される。これらのLSP係数と先行フレームの最後のサブフレームからのLSP係数の間の線形補間は、従来通り、LSP補間プロセッサ350により最初の4つのサブフレームのためのLSP係数を提供するために実行される。このとき、補間され量子化されたLSP係数は、従来の要領でLSP-LPC変換プロセッサ35

5による各々のサブフレーム内での使用のためLPC予測器係数に変換し戻される。これは、符号器及び復号器の両方において行なわれる。LSP補間は、出力音声のスムーズな再生を維持する上で重要である。LSP補間は、サブフレーム(4ms)につき一回ずつスムーズにLPC予測器係数を更新することを可能にする。結果として得られたLPC予測器係数アレイaは、符号器の入力信号を予測する

べくLPC予測誤差フィルタ20の中で使用される。入力信号とその予測されたバージョンの間の差は、LPC予測残余dである。

2. 整形フィルタ

整形フィルタ係数プロセッサ30は、LPC予測器係数アレイaの最初の3つの自己相関係数を計算し、次に、対応する最適な2次全極予測器について c_j 、 $j = 0, 1, 2$ として、係数 C_j を解くため、レヴィンソン-ダービン巡回形を使用する。これらの予測器係数はこのとき0.7という係数で帯域幅拡張される(すなわち、 j 番目の係数 C_j は $C_j (0.7)^j$ で置換される)。次に、プロセッサ30は同様に、16次全極LPC予測器係数アレイaの帯域幅拡張も実行するが、この場合係数は0.8である。これらの2つの帯域幅拡張された全極フィルタ(2次及び16次)をカスケード化することによって、望ましい18次整形フィルタ40が得られる。整形フィルタ係数アレイawcは、直接形の18次フィルタを得るべく、上述の2つの帯域幅拡張された係数アレイ(2次及び16次)を畳み込むことによって計算される。

整形フィルタ40が、図1に示されている通りにLPC予測誤差フィルタでカスケード化された時点で、2つのフィルタは実際に、望まれる符号化雑音スペクトルのほぼ逆数である周波数応答をもつ知覚重みづけフィルタを形成する。かくして、整形フィルタ40の出力は、知覚重みづけ音声信号swと呼ばれる。

ゼロ入力応答プロセッサ50は、その中に整形フィルタを有する。各々の4msのサブフレームの始めに、このプロセッサは、4msのゼロ相当入力信号をフィルタに供給することにより整形ろ過を実行する。一般に、対応する出力信号ベクトルzirは、フィルタが一般に非ゼロメモリを有することから(符号器初期

化の後の一一番最初のサブフレーム中、又は符号器が始動したために符号器の入力信号が正確にゼロであるときを除く）、非ゼロである。プロセッサ60は、重みづけされた音声ベクトル s_w から $z_i r$ を減算する。結果として得られたベクトル t_p は、閉ループピッチ予測のための標的ベクトルである。

3. 閉ループピッチ予測

ピッチ予測において、量子化され復号器に伝送される必要のあるパラメータの種類には2つある。すなわち、有声音声のほぼ周期的な波形の周期に対応するピ

ッヂ周期、及び3つの予測係数（タップ）である。

a. ピッチ周期

LPC予測残余のピッチ周期は、ここにあたかも完全に記述されているごとくに参考として内含されている「有声メッセージ符号器／復号器の使用方法」という題の米国特許第5,327,620号の中で論述された効率の良い2段階探索技術の修正バージョンを用いて、開ループピッチ抽出器及び補間回路70によって、決定される。プロセッサ70は、帯域幅を約700Hzに制限するべく3次楕円低域フィルタの中にLPC残余を通過させ、次に低域フィルタ出力の8:1のデシメーションを実行する。現フレームの最後の3つのサブフレームに対応するピッチ解析窓を用いて、デシメートされた信号の相関係数は、デシメートされていない信号領域内の24~272サンプルのタイムラグに対応する3~34のタイムラグについて計算される。かくして、ピッチ周期として許容可能な範囲は1.5ms~17msあるか又はピッチ周波数でいうと59Hz~667Hzである。これは、低ピッチの男性及び高ピッチの子供を含め大部分の話者の通常のピッチ範囲を網羅するのに充分である。

デシメートされた信号の相関係数が計算された後、最低のタイムラグを有する相関係数の最初の大きいピークが識別される。これが第1段の探索である。結果として得られたタイムラグを t とする。この値 t に8を乗じてデシメートされない信号領域内のタイムラグを得る。結果として得られたタイムラグ $8t$ は、真のピッチ周期がある可能性の最も高い近傍を指している。デシメートされていない信号領域内にもとの時間分解能を保持するため、 $t-4$ から $t+4$ の範囲内で第

2段のピッチ探索が行なわれる。もとのデシメートされていないLPC残余の相関係数dは、 $t - 4 \sim t + 4$ のタイムラグについて計算される（下界は24サンプル、上界は272サンプル）。この範囲内の最大相関係数に対応するタイムラグは、このとき最終ピッチ周期として識別される。このピッチ周期は、8ビットに符号化され、8ビットの指標IPが、サイド情報として復号器に伝送されるためにマルチプレクサ180に供給される。ピッチ周期として選択され得ると考えられる整数は $272 - 24 + 1 = 249$ しかないことから、ピッチ周期を表わすのに8つのビットで充分である。

各々の20msのフレームについてこのような8ビットのピッチ指標は1つずつしか伝送されない。プロセッサ70は、以下の要領で各々のサブフレームについてピッチ周期 k_{pi} を決定する。現フレームの抽出されたピッチ周期と最後のフレームのものの間の差が20%以上である場合、上述の抽出されたピッチ周期は、現フレーム内の全てのサブフレームについて用いられる。一方、この相対的ピッチ変化が20%未満である場合、抽出されたピッチ周期は、現フレームの最後の3つのサブフレームについて使用され、一方最初の2つのサブフレームのピッチ周期は、最後のフレームの抽出されたピッチ周期と現フレームのものの間の線形補間によって得られる。

b. ピッチ予測器タップ

閉ループピッチタップ量子化器とピッチ予測器80は、以下のオペレーションをサブフレーム毎に実行する：すなわち（1）3ピッチタップの閉ループ量子化、（2）現フレーム内のLPC予測残余dのピッチ予測されたバージョンである d_h の生成及び、（3）標的信号 t_p に最も近い突合せである h_d の生成である。

プロセッサ80は、LPC予測残余dの量子化バージョンとみなすことのできる信号 d_t の先行サンプルを記憶する内部バッファを有する。各々のサブフレームについて、プロセッサ80は、 d_t バッファから64次元ベクトルを3つ抽出するため、ピッチ周期 k_{pi} を用いる。 x_1 、 x_2 及び x_3 と呼ばれるこれらの3つのベクトルはそれぞれ、 d_t の現フレームよりも $k_{pi} - 1$ 、 k_{pi} 及び $k_{pi} + 1$ のサンプルを含む。

$i + 1$ サンプルだけ早い。このとき、これらの 3 つのベクトルは、ゼロ初期フィルタメモリをもつ整形フィルタ（係数アレイ $a w c$ を伴う）によって別々にろ過される。結果として得られる 3 つの 64 次元出力ベクトルを y_1 , y_2 及び y_3 と呼ぼう。次に、プロセッサ 80 は、3 ピッチ予測器タップの 64 の候補セット b_{1j} , b_{2j} , b_{3j} ($j = 1, 2, \dots, 64$) のコードブックを通して探索し、ひずみ測定を最小限にする最適なセット b_{1k} , b_{2k} , b_{3k} を見い出す必要がある。

$$\| t p - \sum_{i=1}^3 b_{ik} y_i \|^2$$

このタイプの問題は、以前に研究されており、米国特許第 5,327,520 号

の中に効率の良い探索方法に見い出すことができる。この技術の詳細はここでは紹介しないが、基本的な考え方は以下の通りである。

このひずみ測定を最小にすることが 2 つの 9 次元ベクトルの内部積を最大にすることと同等であるということを示すことができる。これらの 9 次元ベクトルの 1 つは、 y_1 , y_2 及び y_3 の相関係数のみを含んでいる。もう 1 つの 9 次元ベクトルは、評価中の 3 つのピッチ予測器タップのセットから導出された積の項のみを含んでいる。このようなベクトルは信号に依存せず、ピッチタップコードベクトルのみに依存していることから、このようなベクトルの可能性は 64 しかなく（各ピッチタップコードベクトルについて 1 つずつ）、これらを、予め計算し、1 つのテーブルつまり VQ コードブックの中に記憶させることができる。実際のコードブック探索においては、 y_1 , y_2 , 及び y_3 の 9 次元相関ベクトルが最初に計算される。次に、64 の予め計算され記憶された 9 次元ベクトルの各々との結果として得られたベクトルの内部積が計算される。最大内部積を与える記憶されたテーブル内のベクトルが勝者であり、3 つの量子化されたピッチ予測器タップがそこから導出される。記憶されたテーブル内には 64 のベクトルが存在することから、3 つの量子化されたピッチ予測器タップを表わすには、 m 番目のサブフレームのための 6 ビットの指標 IT (m) で充分である。各フレーム内には 5 つのサブフレームが存在することから、全てのサブフレームに用いられる 3 つのピッチタップを表わすためには、1 フレームにつき合計 30 ビットが用いられる。

。これらの30ビットは、サイド情報としてデコーダに対して伝送するためマルチプレクサ180に供給される。

各々のサブフレームについて、上述のコードブック探索方法により3つのピッチタップ b_{1k} , b_{2k} , b_{3k} の最適なセットが選択された後、dのピッチ予測バージョンが、次のように計算される。

$$d \cdot h = \sum_{i=1}^3 b_{ik} \cdot x_i$$

出力信号ベクトル $h \cdot d$ は、以下のように計算される。

$$h \cdot d = \sum_{i=1}^3 b_{ik} \cdot y_i$$

このベクトル $h \cdot d$ は、減算ユニット90によりベクトル $t \cdot p$ から減算される。結果は、変換符号化のための標的ベクトル $t \cdot t$ である。

4. 標的ベクトルの変換符号化

a. 正規化のための整形フィルタ絶対値応答

標的ベクトル $t \cdot t$ は、変換符号化アプローチを用いて、ブロック100-150によりサブフレーム毎に符号化される。整形フィルタ絶対値応答プロセッサ100は、以下の要領で信号 $m \cdot a \cdot g$ を計算する。まず最初に、このプロセッサは、現フレームの最後のサブフレームの整形フィルタ係数アレイ $a \cdot w \cdot c$ をとり、それを64サンプルにゼロパッドし、次に、結果として得られた64次元ベクトルについて64ポイントFFTを実行する。次にこれは、0~8kHzの周波数範囲に対応する33のFFT係数の絶対値を計算する。結果としてのベクトル $m \cdot a \cdot g$ は、最後のサブフレームのための整形フィルタの絶対値応答である。計算を節約するため、最初の4つのサブフレームは、最後のフレームの最後のサブフレームの $m \cdot a \cdot g$ ベクトルと現フレームの最後のサブフレームの $m \cdot a \cdot g$ ベクトルの間の線形補間によって得られる。

b. 変換及び利得正規化

変換プロセッサ110は、以下に記述するようないくつかのオペレーションを実行する。まず第1に、これは、64ポイントFFTを用いることにより現サブ

フレーム内の64次元ベクトルを変換する。64サンプル（又は4ms）というこの変換サイズは、オーディオ符号化技術においては周知のものであるいわゆる「プレ・エコー」ひずみを回避する。ここにあたかも完全に記述されているごとくに参考として内含されている Jayant, N. et al., 「人間の知覚モデルに基づく信号圧縮」Proc. IEEE pp 1385~1422, 1993年10月を参照のこと。最初の33の複合FFT係数の各々は次に、magベクトル内の対応する要素により除算される。結果として得られた正規化されたFFT係数ベクトルは、3つの周波数帯域、すなわち（1）最初の6つの正規化されたFFT係数

から成る低周波数帯域（すなわち0~1250Hz）、（2）次の10個の正規化されたFFT係数から成る中周波数帯域（1500~3750Hz）及び（3）残りの17の正規化されたFFT係数から成る周波数帯域（4000~8000Hz）に区分される。

3つの帯域の各々の中の全エネルギーは計算されて次に、各帯域の対数利得と呼ばれるdB値に変換される。低周波数帯域の対数利得は、当該技術分野において周知のロイドアルゴリズムを用いて設計された5ビットのスカラ量子化器を用いて量子化される。量子化された低周波数対数利得は、中及び高周波数帯域の対数利得から減算される。結果として得られるレベル調整された中及び高周波数の対数利得は2次元ベクトルを形成するべく連結され、これは次に、同じく当該技術分野において周知の一般化されたロイドアルゴリズムにより設計されたコードブックを伴う7ビットベクトル量子化器によって量子化される。量子化された低周波数対数利得は次に、レベル調整された中及び高周波数対数利得の量子化されたバージョンに加算し戻され、中及び高周波数帯域の量子化対数利得が得られる。次に、3つの量子化された対数利得は全て対数（dB）領域から線形領域へ変換される。このとき、33の正規化されたFFT係数（以上で記述したとおりmagによって正規化されたもの）の各々は、さらに、FFT係数がある周波数帯域の対応する量子化された線形利得によって除算される。この正規化第2段の後、結果は、0~8000Hzの周波数を表わす33の複素数を含む最終的な正規化された変換ベクトルt_cである。

m番目のサブフレーム中の対数利得の量子化中、変換プロセッサ110は、低周波数対数利得のための5ビットの利得コードブック指標IG(m, 1)及び、中及び高周波数対数利得のための7ビットコードブック指標IG(m, 2)を生成する。従って、3つの対数利得は、サブフレームあたり12ビットつまりフレームあたり60ビットのビットレートで符号化される。これらの60ビットは、サイド情報として復号器に伝送するためマルチプレクサ180に供給される。これら60の利得ビットは、LSPのための49ビット、ピッチ周期のための8ビット及びピッチタップのための30ビットと共に、サイド情報を形成し、その合計は1フレームあたり $49 + 8 + 30 + 60 = 147$ ビットとなる。

c. ビットストリーム

上述の通り、LPCパラメータを符号化するために49ビット/フレームが割振られ、3-タップピッチ予測器のために $8 + (6 \times 5) = 38$ ビット/フレームが割振られ、利得のために $(5 + 7) \times 5 = 60$ ビット/フレームが割振られた。従って、サイド情報ビットの合計数は20msのフレームあたり $49 + 38 + 60 = 147$ ビットつまり4msのサブフレームあたりおよそ30ビットである。符号器は、16, 24及び32kb/sという3つの異なるレートの1つで使用され得ると考えよう。16kHzのサンプリングルートで、これら3つの標的レートは、1, 1.5及び2ビット/サンプル又は64, 96及び128ビット/サブフレームにそれぞれ変形する。サイド情報のために用いられる30ビット/サブフレームで、主情報を符号化(FFT係数の符号化)する上で使用するべく残っているビットの数は、それぞれ16, 24及び32kb/sの3つのレートについて34, 66及び98ビット/サブフレームである。

d. 適応ビット割振り

本発明の原則に従うと、TCP復号器における出力音声の知覚的品質を高めるため、異なる量子化精度で周波数スペクトルのさまざまな部分にこれらの残りのビットを割当てるべく、適応ビット割振りが実行される。これは、オーディオ信号内の雑音に対する人間の感度のモデルを用いることによって行なわれる。このようなモデルは、知覚的オーディオ符号化技術においては既知のものである。例

えば、Tobias J.V. ed., 「モデル聴覚理論の基礎」、Academic Press New York 及び London, 1970 を参照のこと。同様に、ここにあたかも完全に記述されているかのように参考として内含されている Schroeder, M.R. et al. 「人間の耳のマスキング特性を開発利用することによるデジタル音声符号器の最適化」。J. Acoust. Soc. Amer., 66: 1647-1652, 1979年12月 (Schroeder et al.)も参照のこと。

聴力モデル及び量子化制御プロセッサ 130 が適応ビット割振りを実行し、 t_c 内に含まれた 33 の正規化された変換係数の各々を量子化するためにいくつかのビットを使用すべきかについて変換保数量子化器 120 に告げる出力ベクトル b_a を生成する。適応ビット割振りはサブフレームに 1 回ずつ実行することがで

きるが、本発明の実施形態では、計算上の複雑さを低減するため 1 フレームにつき 1 回ずつビット割振りが行なわれる。

従来の音楽符号器で行なわれているように、雑音マスキングしきい値及びビットの割振りを導出するのに量子化されていない入力信号を使用するのではなく、実施形態の雑音マスキングしきい値及びビットの割振りは、量子化された LPC 合成フィルタ（これは往々にして「LPC スペクトル」と呼ばれる）の周波数応答から決定される。LPC スペクトルは、24ms の LPC 解析窓内の入力信号のスペクトル包絡線の近似とみなすことができる。LPC スペクトルは、量子化された LPC 係数に基づいて決定される。量子化された LPC 係数は、LPC スペクトルを以下のように決定する聴力モデル及び量子化器制御プロセッサ 130 に対して、LPC パラメータプロセッサ 10 によって提供される。量子化された LPC フィルタ係数 a は、まず第 1 に 64 ポイント FFT によって変換される。最初の 33 の FET 係数の各々の累乗が決定され、次に逆数が計算される。結果は、64 ポイント FFT の周波数分解能をもつ LPC パワースペクトルである。

LPC パワースペクトルが決定された後、ここにあたかも完全に記述されているごとくに参考として内含されている米国特許第 5, 314, 457 号の中に記述された方法の修正バージョンを用いて、推定雑音マスキングしきい値 TM が計

算される。プロセッサ130は、主観的リスニング実験から経験的に決定された周波数依存性減衰関数により、LPCパワースペクトルの33のサンプルを基準化する。減衰関数は、LPCパワースペクトルのDC項について12dBから始まり、700～800Hzの間で約15dBまで増大し、次に高周波数に向かって単調に減少し、最終的に8000Hzでの6dBまで減少する。

33の減衰されたLPCパワースペクトルサンプルの各々は、次に、マスキングしきい値を計算するべくその特定の周波数について導出された「基底板広がり関数」を基準化するために用いられる。任意の周波数のための広がり関数は、その周波数における単一音調マスカー信号に応答したマスキングしきい値の形状に対応する。Schroeder et al.の等式(5)は、「バルク」周波数尺度又は臨界帯域周波数の形でこのような広がり関数を記述しており、これはここに、あたかも完全に記述されているかのごとく参考として内含される。尺度化プロセスは、

「バルク」周波数尺度に64ポイントFFTの最初の33個の周波数(すなわち0Hz, 250Hz, 500Hz, … 8000Hz)が変換されることから始まる。次に、結果として得られた33のバーク値の各々について、Schroeder et alの等式(5)を用いてこれらの33のバーク値で、対応する広がり関数がサンプリングされる。結果として得られた33の広がり関数は、1つのテーブルの中に記憶され、これはオフラインプロセスの一部分として行なわれ得る。推定マスキングしきい値を計算するため、33の広がり関数の各々には、減衰されたLPCパワースペクトルの対応するサンプル値が乗じられ、結果として得られた33の基準化された広がり関数は合計される。結果は、推定マスキングしきい値関数である。マスキングしきい値を推定するためのこの技術が、利用可能な唯一の技術ではないという点に留意すべきである。

複雑性を低く保つため、プロセッサ130は適応ビット割振りを実行するに「貧欲」アルゴリズムを使用する。この技術は、将来のビット割振りに対する潜在的な影響とは無関係に、最も「困窮している」周波数成分に対して一度に1つのビットを割振るという意味において「貧欲な」ものである。まだいかなるビットも割当てられていない最初において、対応する出力音声はゼロとなり、符号化

誤り信号は入力音声そのものである。従って当初 L P C パワースペクトルは符号化雑音のパワースペクトルであると仮定される。その後、上で計算されたマスキングしきい値及び Schroeder et al の中で雑音の大きさの計算方法の簡略化されたバージョンを用いて、64 ポイントの F F T の 33 の周波数の各々における雑音の大きさが推定される。

33 の周波数の各々における簡略化された雑音の大きさは、以下のように計算される。まず最初に、Tobias の Scharf の本の章の表 1 に列挙された臨界帯域幅の線形補間を用いて、i 番目の周波数における臨界帯域幅 B_i を計算する。結果は Schroeder et al の等式 (3) の中の項 d_f / d_x の近似値である。33 の臨界帯域幅の値を予め計算し、テーブルの中に記憶する。次に、i 番目の周波数について、雑音パワー N_i をマスキングしきい値 M_i と比較する。 $N_i \leq M_i$ である場合、雑音の大きさ L_i はゼロにセットされる。 $N_i \geq M_i$ である場合、雑音の大きさは、以下のように計算される。

$$L_i = B_i \left((N_i - M_i) / (1 + (S_i / N_i)^2) \right)^{0.25}$$

なお式中、 S_i は i 番目の周波数における L P C パワースペクトルのサンプル値である。

ひとたび雑音の大きさが 33 の周波数全てに計算されたならば、最大の雑音の大きさをもつ周波数が識別され、この周波数に対して 1 つのビットが割当てられる。この周波数における雑音パワーは、正規化された F F T 係数を量子化するため V Q コードブックの設計中に得られた信号雑音比 (S N R) から経験的に決定される 1 つの係数により縮分させられる (縮分係数の値の例は 4 ~ 5 d B である)。この周波数における雑音の大きさは、次に縮分された雑音パワーを用いて更新される。次に、更新された雑音の大きさアレイから最大値が識別され、1 つのビットが対応する周波数に割当てられる。このプロセスは、全ての利用可能ビットが使い尽されるまで続く。

32 及び 24 k b / 秒の T P C 符号器については、33 の周波数の各々が、適応ビット割振りの間にビットを受理することができる。一方、16 k b / 秒の T P C 符号器については、符号器が 0 ~ 4 k H z の周波数範囲のみに対してビット

を割当て（すなわち最初の 16 の F E T 係数）、高周波数合成プロセッサ 140 を用いて 4 ~ 8 k H z のより高い周波数帯域内で残余 F F T 係数を合成する場合に、より優れた音声品質を達成することができる。

量子化された L P C 係数 a は、T P C 復号器でも利用可能であることから、ビット割振り情報を伝送する必要は全くなかった、ということに留意されたい。このビット割振り情報は、復号器内の聴力モデル量子化制御プロセッサ 50 のレプリカにより決定される。かくして、T P C 復号器は、このようなビット割振り情報を得るため符号器の適応ビット割振りオペレーションを局所的に複製することができる。

e. 変換係数の量子化

変換係数量子化器 120 は、ビット割振り信号 b_a を用いて t_c 内に含まれた変換係数を量子化する。F F T の D C 項は実数であり、ビット割振りの間に何らかのビットを受理した場合、スカラー量子化される。それが受理できるビットの最大数は 4 である。2 番目から 16 番目までの F F T 係数については、実数部分

及び虚数部分を同時に量子化するために、従来の 2 次元ベクトル量子化器が使用される。この 2 次元 V Q のためのビットの最大数は 6 ビットである。残りの F F T 係数については、2 つの隣接する F F T 係数の実及び虚部分を同時に量子化するために、従来の 4 次元ベクトル量子化器が使用される。変換係数の量子化の後、結果として得られる V Q コードブック指標アレイ I C は、T P C 符号器の主情報を含む。この指標アレイ I C は、マルチプレクサ 180 に供給され、ここでサイト情報ビットと組合わされる。結果は、通信チャンネルを通って T P C 復号器に伝送される最終ビットフレームである。

変換係数量子化器 120 は同様に、正規化された変換係数の量子化された値を復号する。これは次に、対応する m a g 要素及び対応する周波数帯域の量子化された線形利得をこれら係数の各々に乘じることにより、これらの変換係数のもとの利得レベルを復元する。結果は出力ベクトル $d_t c$ である。

f. 高周波数合成及び雑音充てん

16 k b / 秒の符号器については、適応ビット割振りは 0 ~ 4 k H z の帯域に

制限され、プロセッサ140は4～8 kHzの帯域を合成する。それを行なう前に、聴力モデル量子化器制御プロセッサ130はまず第1に、4～7 kHzの帯域内の周波数について、LPCパワースペクトルとマスキングしきい値の間の比率又は信号-マスキングしきい値比(SMR)を計算する。17番目～29番目のFFT係数(4～7 kHz)は、ランダムである位相及びSMRによって制御されている絶対値を用いて合成される。SMR > 5 dBである周波数については、FFT係数の絶対値は、高周波数帯域の量子化された線形利得にセットされる。SMR ≤ 5 dBである周波数については、絶対値は、高周波数帯域の量子化された線形利得より2 dB下である。30番目から33番目のFET係数については、絶対値は、高周波数帯域の量子化された線形利得よりも2 dB～30 dBまで傾斜し、位相は再びランダムである。

32 kb/s及び24 kb/sの符号器については、記述されている通り、周波数帯域全体について、ビット割振りが実行される。しかしながら、4～8 kHzの帯域内のいくつかの周波数はなおも全くビットを受理しない可能性がある。この場合、上述の高周波数合成及び雑音充てん手順は、ビットを全く受理しない

周波数のみに適用される。

ベクトル $d_t c$ に対してこのような高周波数合成及び雑音充てんを適用した後、結果として得られる出力ベクトル $q_t c$ は、正規化の前の変換係数の量子化バージョンを含む。

g. 逆変換及びフィルタメモリ更新

逆変換プロセッサ150は、半サイズの33要素ベクトル $q_t c$ によって表わされる64要素複合ベクトルについて逆FFTを実行する。この結果として、変換コーディングのための時間-領域標的ベクトル t_t の量子化されたバージョンである出力ベクトル $q_t t$ が得られる。

ゼロ初期フィルタ状態(フィルタメモリ)では、その係数アレイとして $a_w c$ をもつ全ゼロフィルタである逆整形フィルタ160は、出力ベクトル e_t を生成するべくベクトル $q_t t$ をろ過する。このとき、加算器170は d_h を e_t に加算して、量子化されたLPC予測残余 d_t を得る。この d_t ベクトルは次に、閉

ループピッチタップ量子化器及びピッチ予測器 80 の中の内部記憶バッファを更新するために用いられる。これは又、次のサブフレームのためのゼロ入力応答生成に備えて補正フィルタメモリを樹立するためゼロ入力応答プロセッサ 50 の内部の内部整形フィルタを励起するためにも使用される。

C. 復号の実施形態

本発明の復号実施形態が、図 2 に示されている。各々のフレームについて、デマルチプレクサ 200 は、受理されたビットストリームから全ての主及びサイド情報成分を分離する。主情報、つまり変換係数指標アレイ IC は、変換係数復号器 235 に供給される。この主情報を復号するためには、いくつの主情報ビットが各々の量子化された変換係数と結びつけられるかを決定するため、適応ビット割振りを実行しなければならない。

適応ビット割振りにおける第 1 のステップは、量子化された LPC 係数（割振りを左右するもの）の生成である。デマルチプレクサ 200 は、LPC パラメータ復号器 215 に対して 7 つの LSP コートブック指標 IL (1) ~ IL (7) を供給し、この復号器は、16 の量子化された LSP 係数を得るために 7 つの LSPVQ コードブックからテーブルのルックアップを実行する。このとき、LPC

パラメータ復号器 215 は、図 3 のブロック 345, 350 及び 355 と同じインテイニング、補間及び LSP - LPC 係数変換オペレーションを実行する。

計算された LPC 係数アレイ a で、聴力モデル量子化器制御プロセッサ 220 は、TPC 符号器内のプロセッサ 130 と同じ要領で、各々の FFT 係数について（量子化された LPC パラメータに基づいて）ビット割振りを決定する（図 1）。同様にして、整形フィルタ係数プロセッサ 225 及び整形フィルタ絶対値応答プロセッサ 230 も、TPC 符号器の中でそれぞれ対応するプロセッサ 30 及び 100 のレプリカである。プロセッサ 230 は、変換係数復号器 235 が使用するように、整形フィルタの絶対値応答 mag を生成する。

ひとたびビット割振り情報が導出された時点で、変換係数復号器 235 は次に主情報を正しく復号し、正規化された変換係数の量子化された情報を得ることができる。復号器 235 は同様に、利得指標アレイ IG を用いて利得を復号する。

各々のサブフレームについて、2つの利得指標（5ビット及び7ビット）が存在し、これらは、低周波数帯域の量子化された対数利得及び、中及び高周波数対数利得のレベル調整された対数利得の量子化バージョンへと復号される。このとき、量子化された低周波数対数利得は、中及び高周波数帯域の量子化された対数利得を得るべく、レベル調整された中及び高周波数対数利得の量子化バージョンに加算し戻される。次に3つの量子化された対数利得は全て、対数（dB）領域から線形領域まで変換される。3つの量子化された線形利得の各々は、対応する周波数帯域内の正規化された変換係数の量子化されたバージョンを乗算するのに用いられる。結果として得られた3つの利得基準化された量子化変換係数の各々は、次に、整形フィルタ絶対値応答アレイmagの中の対応する要素でさらに乗算される。これらの2つの基準化段の後、結果は、復号された変換係数アレイdtcである。

高周波数合成プロセッサ240、逆変換プロセッサ245及び逆整形フィルタ250は、ここでもTPC符号器内の対応するブロック（140, 150及び160）の正確なレプリカである。これらは、合わせて、高周波数合成、雑音充てん、逆変換、及び逆整形ろ過を実行して、量子化された励起ベクトルetを生成する。

ピッチ復号器及び補間回路205は、最後の3つのサブフレームのためのピッチ周期を得るために8ビットピッチ指標IPを復号し、次に、TPC符号器の対応するブロック70内で行なわれたのと同じ要領で最初の2つのサブフレームのためのピッチ周期を補間する。ピッチタップ復号器及びピッチ予測器210は、3つの量子化されたピッチ予測器タップ b_{1k}, b_{2k}, b_{3k} を得るために、各々のサブフレームについてピッチタップ指標ITを復号する。次に、これは、補間されたピッチ周期 k_{pi} を用いて、符号器の節で記述したものと同じ3つのベクトル x_1, x_2, x_3 を抽出する。（これら3つのベクトルはそれぞれ、dtの現フレームよりも $k_{pi}-1, k_{pi}$ 及び $k_{pi}+1$ サンプルだけ早い）。次にこれは以下のとおり、LPC残余のピッチ予測されたバージョンを計算する：

$$d \cdot h = \sum_{i=1}^3 b_{ik} \cdot x_i$$

加算器255は、 $d \cdot h$ を $e \cdot t$ に加算して、LPC予測残余 d の量子化バージョン $d \cdot t$ を得る。この $d \cdot t$ ベクトルは、 $d \cdot t$ についてその内部記憶バッファを更新する（ピッチ予測器のフィルタメモリ）ために、ブロック210内のピッチ予測器までフィードバックされる。

長期後置フィルタ260は、ITU-T G.728の標準的16kb/sの低遅延CELP符号器の中で用いられる長期後置フィルタと基本的に類似している。主たる相異点は、有声化インジケータとして3つの量子化されたピッチタップの合計である

$$\sum_{i=1}^3 b_{ik}$$

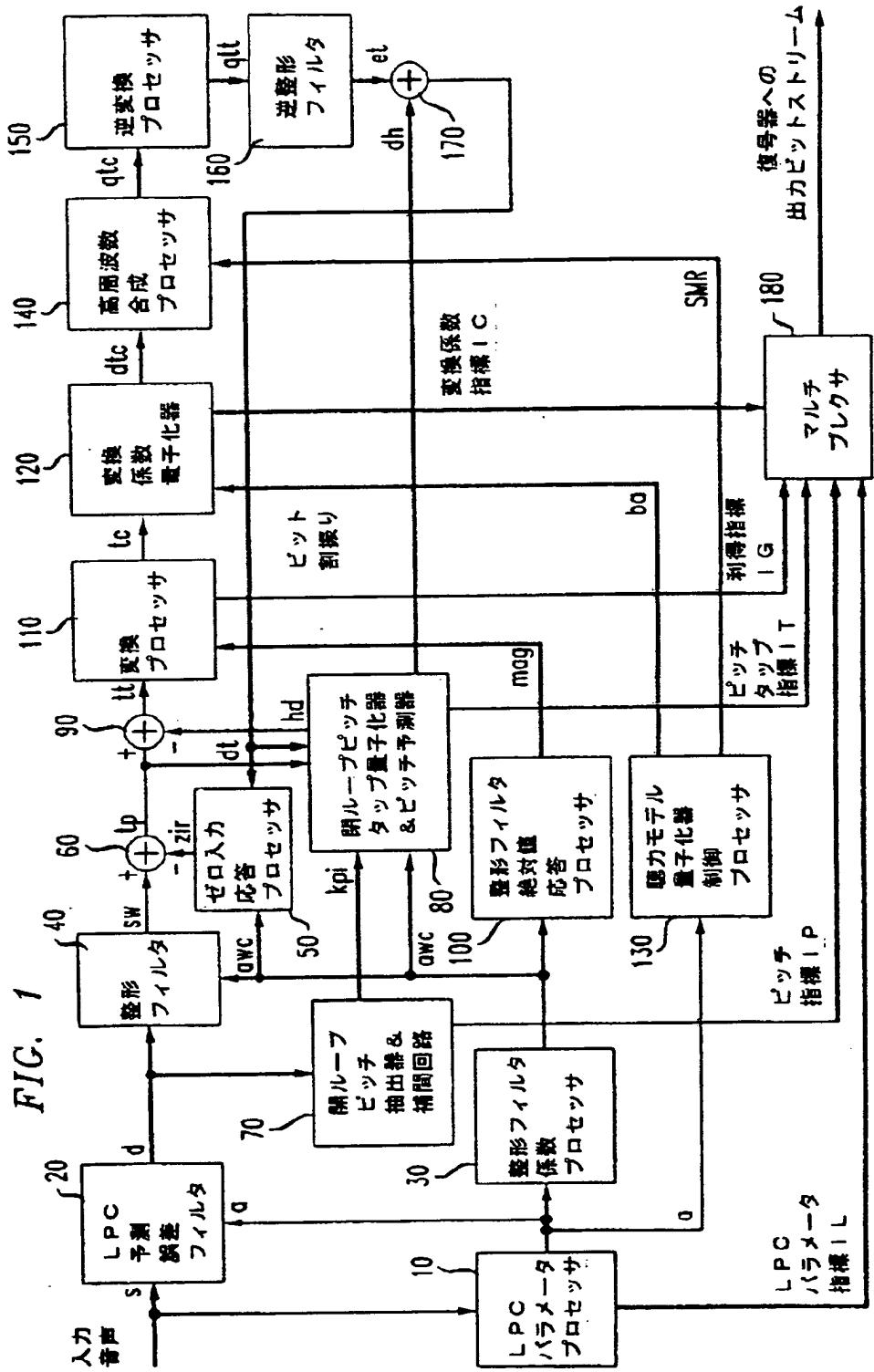
を用いること、そして長期後置フィルタ係数のための基準化因子がG.728にあるように0.15ではなく0.4であること、にある。この有声化インジケータが0.5未満である場合、事後ろ過オペレーションは省かれ、出力ベクトル $f \cdot d \cdot t$ は入力ベクトル $d \cdot t$ と同一である。インジケータが0.5以上である場合、事後ろ過オペレーションが実施される。

LPC合成フィルタ265は、標準的なLPCフィルタ、すなわち量子化されたLPC係数アレイ a を伴う全極直接形フィルタである。これは、信号 $f \cdot d \cdot t$ をろ過し、長期後置ろ過された量子化音声ベクトル $s \cdot t$ を生成する。この $s \cdot t$ ベクトルは、短期後置フィルタ270の中を通過させられて、最終的TPC復号器出力音声信号 $f \cdot s \cdot t$ を生成する。ここでも又、この短期後置フィルタ270は、G.728の中で使用される短期後置フィルタに非常に類似している。唯一の差異は、以下の点にある。まず第1に、極制御係数、ゼロ制御係数及びスペクトル傾動制御係数がそれぞれ、G.728における0.75, 0.65及び0.15という対応値ではなく、0.7, 0.55及び0.4である。第2に一次スペクトル傾動補償フィルタの係数は、フレーム間でサンプル毎に線形補間される。こうして、フレーム境界における不連続性に起因する、場合によっては可聴となるクリック

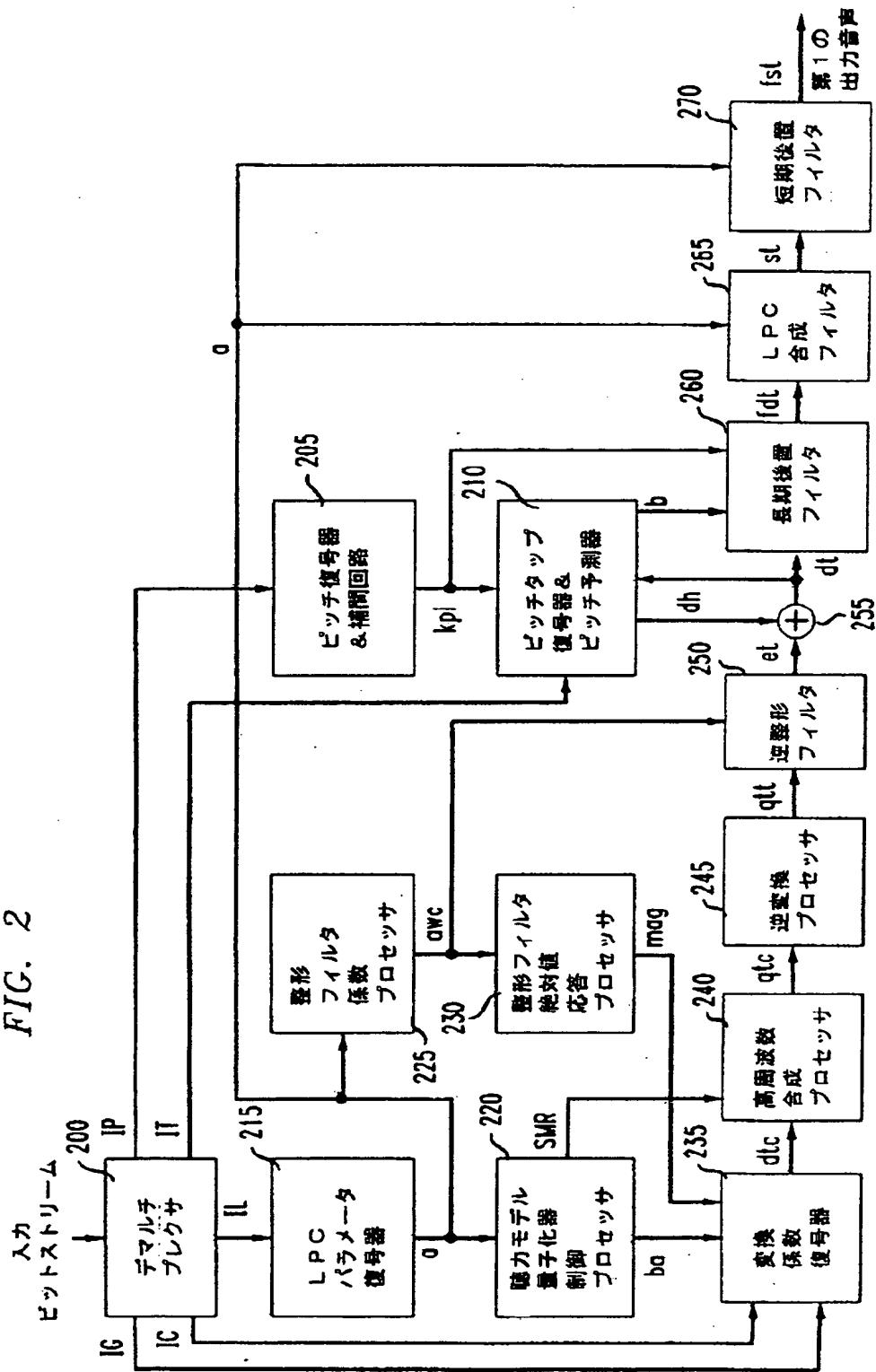
音が避けられる。

長期及び短期後置フィルタは、出力信号 $f_s t$ において符号化雑音の知覚されたレベルを低下させ、かくして音声品質を高める効果をもつ。

【図 1】

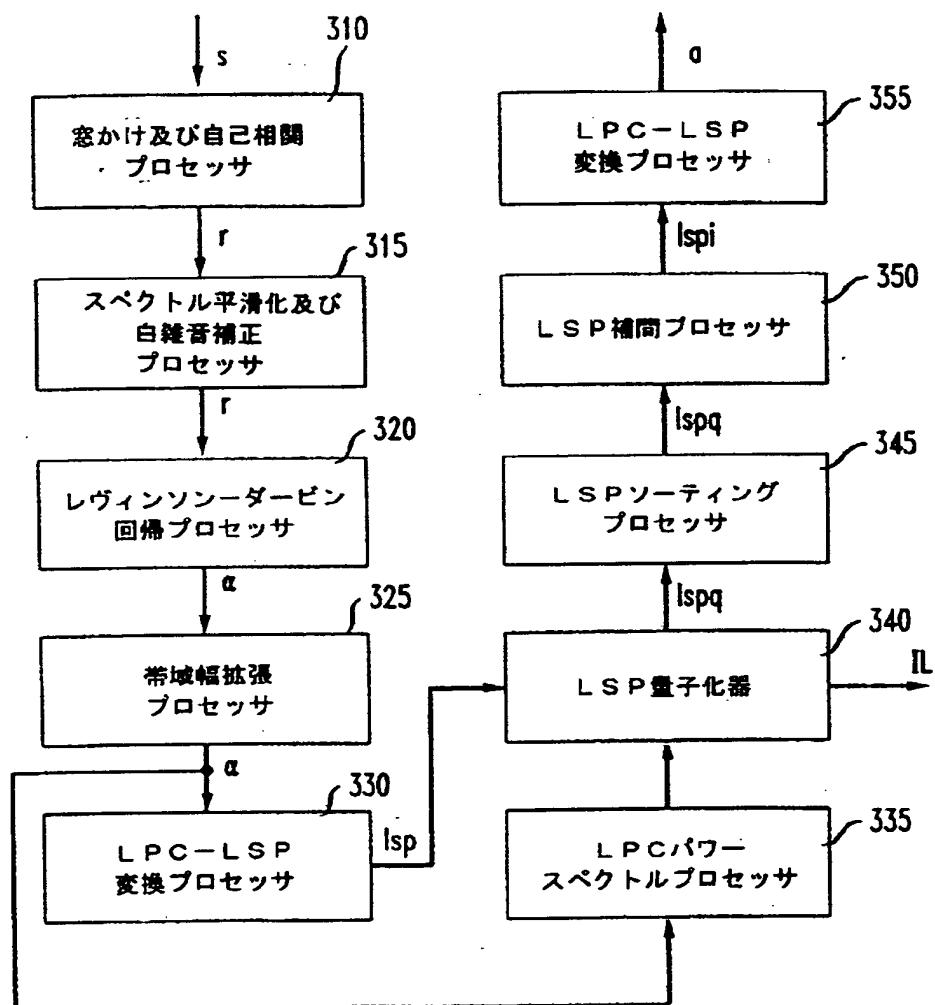


【図2】



【図3】

FIG. 3



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No. PCT/US97/02898

A. CLASSIFICATION OF SUBJECT MATTER IPC(6) : G10L 3/02, 9/00 US CL : 395/2.28, 2.1, 2.12 According to International Patent Classification (IPC) or to both national classification and IPC													
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S. : 395/2.28, 2.1, 2.12													
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched US CL 395/2.09-2.95													
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) APS; IEEE AND IEE CDROM DATABASE; SMART PATENT CDROM DATABASE													
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 2px;">Category*</th> <th style="text-align: left; padding: 2px;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="text-align: left; padding: 2px;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="text-align: center; padding: 2px;">X,P</td> <td style="padding: 2px;">US 5,583,963 A (LOZACH) 10 December 1996 (10/12/96), column 4 line 64 - column 8 line 23, column 12 line 22-67.</td> <td style="text-align: center; padding: 2px;">1</td> </tr> <tr> <td style="text-align: center; padding: 2px;">X</td> <td style="padding: 2px;">US 5,012,517 A (WILSON ET AL) 30 April 1991 (30/04/91), column 7 line 12 - column 13 line 65.</td> <td style="text-align: center; padding: 2px;">1</td> </tr> <tr> <td style="text-align: center; padding: 2px;">Y</td> <td style="padding: 2px;">JOHNSON et al.; "Pitch-Orthogonal Code-Excited LPC.", Global Telecommunications Conference, IEEE GLOBECOM 90, pages 542-546.</td> <td style="text-align: center; padding: 2px;">1</td> </tr> </tbody> </table>		Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X,P	US 5,583,963 A (LOZACH) 10 December 1996 (10/12/96), column 4 line 64 - column 8 line 23, column 12 line 22-67.	1	X	US 5,012,517 A (WILSON ET AL) 30 April 1991 (30/04/91), column 7 line 12 - column 13 line 65.	1	Y	JOHNSON et al.; "Pitch-Orthogonal Code-Excited LPC.", Global Telecommunications Conference, IEEE GLOBECOM 90, pages 542-546.	1
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.											
X,P	US 5,583,963 A (LOZACH) 10 December 1996 (10/12/96), column 4 line 64 - column 8 line 23, column 12 line 22-67.	1											
X	US 5,012,517 A (WILSON ET AL) 30 April 1991 (30/04/91), column 7 line 12 - column 13 line 65.	1											
Y	JOHNSON et al.; "Pitch-Orthogonal Code-Excited LPC.", Global Telecommunications Conference, IEEE GLOBECOM 90, pages 542-546.	1											
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.													
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be part of particular relevance "E" earlier document published no or after the international filing date "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reasons (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" documents published prior to the international filing date but later than the priority date claimed													
Date of the actual completion of the international search 28 APRIL 1997	Date of mailing of the international search report 06.06.1997												
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-9508	Authorized officer  ALLEN MACDONALD Telephone No. (703) 305-9708												

INTERNATIONAL SEARCH REPORT

International application No. PCT/US97/02898

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OFER ET AL.; "A Unified Framework for LPC Excitation Representation in Residual Speech Coders"; Acoustics, Speech & Signal Processing Conference, IEEE ICASSP '89, pages 41-44 especially Section 5 page 44.	1
Y	DAVIDSON G. et al.; "Multiple-Stage Vector Excitation Coding of Speech Waveforms"; Acoustics, Speech & Signal Processing Conference, IEEE ICASSP '88, pages 163-166 especially Section 3 pages 164-166.	1